

# Individual Differences in Exploration and Content Curation Activities within a Cultural Heritage Digital Library

Paula Goodale  
University of Sheffield  
Sheffield  
United Kingdom  
p.goodale@sheffield.ac.uk

## ABSTRACT

This paper presents empirical results from an evaluation study of a cultural heritage digital library. It focuses on the differences in preferences between novice and expert users for functionality supporting browsing and exploration, when engaged in orientation and content curation tasks. Findings indicate both similarities and differences between novice and expert users. Recommendations for future work are proposed.

## Keywords

Novice, expert, information access, information seeking, exploration, content curation, cultural heritage, digital library.

## 1. INTRODUCTION

As digital cultural heritage collections become larger and more widely available, they are targeted at more diverse user communities with varying levels of subject and domain knowledge. No longer the preserve of scholarly researchers, they also seek to engage users with general as well as specialist knowledge, for leisure and education purposes. Users are therefore likely to span across a continuum from novice to expert, with varying interests in the library content, varying degrees of subject and domain knowledge, and different types of task that are likely to be undertaken.

Novice users (low subject and domain knowledge) frequently experience difficulties in finding content via web search and in digital libraries of all kinds, particularly when the task is less focused and more exploratory in nature. Their lack of subject and domain knowledge inhibits the successful use of the search box, as keyword formulation and reformulation often proves difficult. In contrast, expert users, with higher levels of subject and domain knowledge, are more confident in search, as they have a repertoire of topics and associated keywords to draw upon.

It might therefore be expected that novice users will have a preference for tools which support browsing and exploration (discovery) of the digital library content, especially in more diverse and large-scale collections. As digital collections grow, individually and in aggregate forms, simple orientation (understanding 'what's here', i.e. which topics are covered and in what depth) can be challenging, and might need to be addressed even before exploration of the content can begin. Additionally, discovery tools should support the needs of novice users in finding and selecting content for topic-focused tasks. This need is likely to be especially acute when an element of creativity and synthesis is involved, such as content curation.

In Proceedings of 1st International Workshop on Accessing Cultural Heritage at Scale (ACHS'16), 22<sup>nd</sup> June 2016, Newark, NJ, USA. Copyright 2016 for this paper by its authors. Copying permitted for private and academic purposes.

Developers of information seeking support systems that intend to support users in more exploratory and creative tasks, including cultural heritage digital libraries, should therefore seek to provide tools for orientation, finding (non-search), and curating content.

This paper aims to examine these requirements via a laboratory-based evaluation study of an experimental system (PATHS<sup>1</sup>) that offers these types of functionality for a large-scale aggregated cultural heritage digital library, based upon a UK sub-set of the Europeana<sup>2</sup> content. Specifically, the paper aims to investigate any potential differences in the preferences of novice and expert users for these types of tools when engaged in orientation, finding and content curation tasks.

## 2. RELATED WORK

### 2.1 Information seeking tasks and systems in digital cultural heritage

Information seeking tasks in the cultural heritage domain are often more complex and/or exploratory in nature, including subject-based searches and less-focused activities, where there is a higher degree of uncertainty in what is being sought [8][1]. Exploratory information seeking activities go beyond simple look-up or known-item search, incorporating elements of learning (acquiring, interpreting, comparing, etc.) and investigation (analysis, evaluation, synthesis, transformation, etc.) [6]. Information seeking support systems in the area of exploratory search therefore require a wider range of functionality to support these more complex activities [7, 15].

The wider range of user interactions in the cultural heritage domain incorporates content curation and support categories [10]. The second category, curation goes beyond finding into various elements of information use, including the addition of annotations, creation of user exhibitions from available content, and storytelling [10]. These activities are more closely aligned with information use than with information finding (searching, browsing and exploration), and represent an opportunity for cultural heritage digital libraries to provide wider access to content and to support reuse and creativity.

Another important element of user requirements in digital cultural heritage is visual representation of collection items [9]. Support for serendipity can also prove to be beneficial and popular with users engaged in less-focused information seeking tasks [12].

### 2.2 Novice and expert user differences

Differences in the needs and behaviors of novice and expert information seekers has been researched in many domains. In web

---

<sup>1</sup> PATHS Project: <http://www.paths-proejct.eu>

<sup>2</sup> Europeana: <http://www.europeana.eu/portal/>

search, domain expertise results in different search strategies and more successful results in finding relevant content [13]. Domain knowledge also results in more focused, systematic search tactics within digital libraries [14]. However, whilst domain knowledge enhances search success, technical skills may offset this to some degree, thereby indicating that those lacking in both domain knowledge and web search expertise are doubly disadvantaged [4].

In the cultural heritage domain, more experienced users are likely to be scholars and researchers in humanities subject areas, as well as cultural heritage professionals, whilst less experienced users may be from educational and general interest categories [3, 11]. Expert users in cultural heritage undertake a wide variety of tasks including known-item search and more exploratory activities [1]. Moreover, novice users involved in leisure activities also undertake a variety of information seeking tasks, and are highly visually focused, as well as engaging in elements of meaning-making [9].

### 3. METHODS

The results presented in this paper are derived from a comprehensive evaluation study of a prototype of an information seeking support system designed to investigate functionality for the support of exploration and curation of content in large-scale cultural heritage digital libraries, created during the PATHS project. The study was carried out under controlled conditions in a laboratory setting, utilizing a variety of simulated work tasks [2] as a means of gaining feedback on system usability and usefulness, to inform future system design, and to investigate user preferences, behaviors, and interactions in this relatively novel context. Screenshots of the system are shown in figures 1-3 below, illustrating thesaurus, map and path functionality, offered as different means of exploring the content in the collection and of curating content. The prototype PATHS system contained c.1 million items selected from UK institutions in the Europeana digital library,

#### 3.1 Tasks

During the evaluation session users were invited to complete five short orientation and information seeking tasks lasting 5 minutes each, followed by one 30-minute content curation task. This paper focuses on the results of one of the orientation tasks and the content curation task.

The orientation task required users to investigate the topics available in the collection, using any of three tools designed to support browsing and exploration (thesaurus, tag cloud and map). Feedback was then supplied on the ease of use and usefulness of each tool using 5-point semantic differential scales, and the user's rank order of preference for the three tools (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>).

The content curation task entailed finding and selecting content (items held within the digital library) on a topic of the user's own choice, then organizing and annotating these items to form a meaningful route (path) through the collection. This task therefore required the user to employ tactics to find content via the search box and/or the exploration tools used in the earlier orientation task, as well as the more creative element of the activity. The whole task can be considered as exploratory [5] as it is relatively non-prescriptive and open-ended, and incorporates elements of discovery and synthesis [6].

### 3.2 Sample

Sample size was 34 participants, comprising 24 novice users and 10 expert users. Novice users were categorized as those with a more general knowledge of cultural heritage (low subject/domain knowledge), and expert users as those with a higher degree of subject knowledge gained from accessing cultural heritage collections for work-related use. A majority of users (n=32) self-reported either an intermediate or high level of experience in using web search, which it has been suggested may offset a lack of subject and domain knowledge to some degree [4].

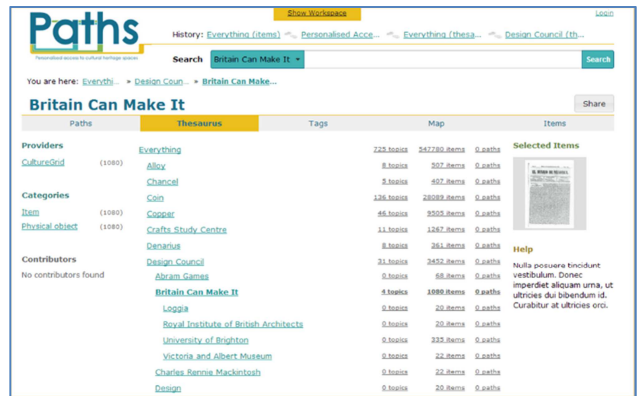


Figure 1: PATHS Screenshot – thesaurus exploration



Figure 2: PATHS Screenshot – map exploration

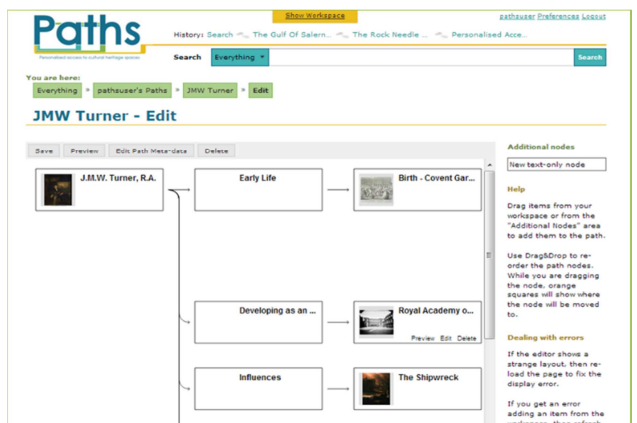


Figure 3: PATHS Screenshot – path creation interface

## 4. RESULTS

Data from user feedback on the two tasks was analysed for user differences according to the novice and expert categorization.

### 4.1 Orientation

Both novice (66.7%) and expert (70%) user types were emphatic in their placement of the thesaurus as the most useful for aiding orientation, i.e. finding out ‘what’s here’ (Table 1). There was more of a split for the tag cloud and the map, with a majority of novice users placing each of these in 3<sup>rd</sup> place, whilst expert users placed these more emphatically in 2<sup>nd</sup> and 3<sup>rd</sup> places respectively. A majority of both user types placed the relatively novel ‘map’ tool in third place, although more of each type also placed it in first position than they did the tag cloud. This difference may be accounted for by the relative novelty of the map, but other factors may also be at play, such as a preference for image vs text visualizations.

		Thesaurus	Tag cloud	Map
Novice	1st	<b>66.7%</b>	12.5%	20.8%
	2nd	33.3%	41.7%	25.0%
	3rd	0.0%	<b>45.8%</b>	<b>54.2%</b>
Expert	1st	<b>70.0%</b>	0.0%	30.0%
	2nd	10.0%	<b>80.0%</b>	10.0%
	3rd	20.0%	20.0%	<b>60.0%</b>

**Table 1: Preference for exploration tools, novice/expert users**

Similarly, 79% of novice users and 80% of expert users rated the thesaurus as either very useful or useful, and 75% and 90% respectively rated it as very easy or easy to use, on 5-point semantic differential scales. However, a difference of opinion was found on the tag cloud, with novice users rating it as less useful (33%) and easy to use (50%), than expert users (80% each useful and easy to use). In contrast, novices were somewhat more favorable towards the map tool, 46% useful and easy to use, than expert users, 40% useful and easy to use.

It seems therefore that the thesaurus is the overall winner for both user types, but that novice users found the map more useful than the tag cloud, and vice versa for expert users.

### 4.2 Finding content

Feedback on the usefulness of tools in finding content of interest for the content curation task was given on a wider range of functionality, including the search box, the thesaurus, tag cloud and map tools, browsing of search results and filtering using facets, recommendations in the form of selected (featured) and related items, metadata, and links to background information in Wikipedia. Again a 5-point differential scale was used (very useful to useless), with an additional category for ‘did not use’.

As might be expected, all users used the search box, although expert users were more emphatic in it being very useful (80%) than novice users (66.7%). As in the orientation task, the thesaurus was deemed the most useful exploration tool, with 46% of novices finding it very useful or useful, compared with 20% of expert users.

Expert users were more likely to rate the usefulness of metadata driven tools, including facets (40%) and metadata keyword links (80%) than novice users (25% and 42% respectively). Interestingly, experts were also more likely to find useful the recommendations in the form of related and selected items, and browsing of search results pages, than novice users. This

unexpected finding for search results pages may arise from more successful searches by expert users, or simply that they had a better idea of what they were looking for and would ‘know it when I see it’.

Overall then, it seems that novices rate the thesaurus most highly of all the exploratory tools offered, and that experts are more likely to find a wider range of tools useful, including those such as facets and subject metadata that might require more specialist knowledge to interpret.

### 4.3 Curating content

The first stage of curating content is to select items for inclusion. Whilst directly related to finding content, there is a more active level of intellectual effort, with choices being made amongst available content, and potentially disregarding some items in favor of others. Users gave feedback on both the information used to make these decisions and the criteria by which items were selected.

As expected, all users, novice and expert, favored images as a primary element of their decision-making process (Table 2). This is unsurprising since it is widely accepted that using cultural heritage collections is a highly visual process, and the curatorial task may be even more visual in nature. It is also clear that novices used much less ‘other’ non-visual information than expert users in making their selections. This difference is most marked in relation to metadata, used by 60% of expert users, but only 12.5% of novice users.

		Novice	Expert
Information used	<i>image</i>	95.8%	100.0%
	<i>title</i>	66.7%	80.0%
	<i>description</i>	50.0%	70.0%
	<i>metadata</i>	12.5%	60.0%
Criteria used	<i>typical</i>	75.0%	40.0%
	<i>unusual/unique</i>	4.2%	10.0%
	<i>aesthetics</i>	62.5%	60.0%
	<i>interesting</i>	29.2%	30.0%
	<i>available</i>	33.3%	30.0%

**Table 2: Information and criteria used for selecting content, novice/expert users**

Criteria used for inclusion of specific items had commonalities and differences (Table 2). Novices and experts were relatively similar in their choice of aesthetically pleasing items (62.5% and 60% respectively), reinforcing the finding on the importance of images. Both user types were similar in their selection based upon interesting descriptions and choosing the only items available on their chosen topic. However, novices (75%) were much more likely to choose typical examples than expert users (40%).

At the next stage of content curation, the items must be arranged in some order and might also be augmented with annotations to add context and aid understanding by the eventual user. There is a striking difference between novice and expert users in ordering their content. Expert users arranged content by theme (40%) and narrative (50%). A majority of novice users also preferred a thematic arrangement (54%), but smaller proportions used criteria such as chronology, geography, narrative, geography, importance, and no particular order. This may indicate that experts have a more specific idea about the nature of curation, incorporating

themes and narratives, but it is also clear that less-experienced users are also drawn towards thematic arrangements.

Finally, novice users were less critical of the curated content they produced during this task. Rating the quality of their output on a scale of 1-10, 21% of novices selected a score of 6 or above, compared to none of the expert users. In contrast, 60% of experts rated their output in the range 1-3, compared with 50% of novices. Additionally, the highest rating given by expert users was 5 out of 10, compared to 9 out of 10 for novice users. It seems that expert users had a clearer idea of what their curated content should look like, both in terms of arrangement and quality of content. In free text feedback, many users commented that they would like better quality images and time to add more contextual annotations to their curated content.

## 5. DISCUSSION

During this study, we have investigated the differences between novice and expert users in their preferences and choices for tools to support more exploratory information seeking and in information use in the form of content curation, within the context of a large-scale aggregated cultural heritage digital library. Whilst search was still the primary choice for all users, novices were more likely to use exploratory tools to augment their orientation and finding activities. Specifically, novices were found to be more pre-disposed to using a thesaurus tool for exploration of the content than expert users, and were also more open to using other exploratory tools. In contrast, experts were more likely to make use of more specialist tools based upon collection metadata, such as facets and subject keywords.

Given the challenges experienced by novices from lower levels of subject and domain knowledge, it is likely that these results may be at least partially explained by the support provided by the exploration tools in overcoming this lower level of knowledge. The thesaurus in particular lays open the main topics within the collection, and is easy to navigate, comprising hierarchical categories and sub-categories. A further bonus may be that the thesaurus was derived from Wikipedia subject headings [ref anon], giving a more informal level of access to subject-related content.

However, differences by novice and expert categorization may not be the only factors affecting accessibility of cultural heritage content. Previous analyses of this evaluation study have also identified differences in behavior and preferences according to cognitive style [ref anon], selected demographics [ref anon] and variations in the system functionality from simple to more complex [ref anon]. It is therefore even more pertinent to consider designing for a diverse range of users to ensure the greatest potential for increasing access, although perhaps focusing on those tools that aid the widest range of users, in this case the thesaurus which was well-received by novices and experts alike.

## 6. CONCLUSIONS

User differences can impact upon successful access to content within large-scale cultural heritage digital libraries. Out of all of these criteria though, it is likely that the novice / expert differences are most likely to affect overall success in finding and exploring content. Novice and expert users express somewhat different preferences for tools to support exploration of digital cultural heritage collections. They also make some different and some similar choices when engaged in finding and creating material for content curation activities. As information seeking support systems for collections are increasingly targeted at a more diverse range of users from novice to expert in their range of

subject and domain knowledge, it is therefore necessary to understand and accommodate these user requirements and differences through functionality that supports a range of preferences and abilities.

In future work we will also undertake more detailed analysis of actual user behavior from screen recordings and transaction logs. This will provide a useful contrast in what users report as preferences and choices, against what functionality they use in practice, as well as uncovering sequences and patterns of behavior, providing a basis for recommendations for system design for the support of exploration in cultural heritage collections. Further, more naturalistic studies of users interacting with systems that are in the public domain, undertaking their own tasks under less controlled conditions will also be of interest, to provide insights into the levels of take-up and actual usage of these types of information seeking support tools in cultural heritage collections 'in the wild'.

## 7. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°. 270082. The author acknowledges the contribution of all project partners involved in PATHS (see: <http://www.paths-project.eu>).

## 8. REFERENCES

- [1] Amin, A. et al. 2008. Understanding Cultural Heritage Experts' Information Seeking Needs. *JCDL'08, June 16-20, 2008, Pittsburgh, Pennsylvania, USA*, 39–47.
- [2] Borlund, P. 2010. Reconsideration of the Simulated Work Task Situation: A Context Instrument for Evaluation of Information Retrieval Interaction. *IiX 2010, August 18-21, 2010, New Brunswick, NJ, USA*, 155–164.
- [3] Goodale, P. et al. 2011. *D 1 . 1 User Requirements Analysis*. PATHS Project <http://www.paths-project.eu/eng/Resources>.
- [4] Hölscher, C. and Strube, G. 2000. Web search behavior of Internet experts and newbies. *Computer Networks*. 33, 1-6, 337–346.
- [5] Kules, B. and Hill, C. 2009. Designing Exploratory Search Tasks for User Studies of Information Seeking Support Systems. *JDCL'09, June 15-19, 2009, Austin, TX, USA*.
- [6] Marchionini, G. 2006. Exploratory Search: From finding to understanding. *Communications of the ACM*. 49, 4, 41–46.
- [7] Shneiderman, B.E.N. 2000. Creating Creativity: User Interfaces for Supporting Innovation. *ACM Transactions of Computer-Human Interaction*. 7, 1, 114–138.
- [8] Skov, M. 2009. *The Reinvented Museum: Exploring Information Seeking Behaviour in a Digital Museum Context*. Mette Skov (thesis). Royal School of Library and Information Science, Denmark.
- [9] Skov, M. and Ingwersen, P. 2008. Exploring Information Seeking Behaviour in a Digital Museum Context. *IiX'08, Information Interaction in Context 2008, London, UK*, 110–115.

- [10] Stiller, J. 2012. A Framework for Classifying Interactions in Cultural Heritage Information Systems. *International Journal of Heritage in the Digital Era*. 1, 1, 141–146.
- [11] Sweetnam, M.S. et al. 2012. User Needs for Enhanced Engagement with Cultural Heritage Collections. *TPDL 2012, Sept 23-27, 2012, Paphos, Cyprus*, 64–75.
- [12] Toms, E.G. and Mccay-peet, L. 2009. Chance Encounters in the Digital Library. *ECDL 2009, Sept 27-Oct 02, 2009, Corfu, Greece*, 192–202.
- [13] White, R.W. et al. 2009. Characterizing the Influence of Domain Expertise on Web Search Behavior. *WSDM'09, February 9-12 2009, Barcelona, Spain*.
- [14] Wildemuth, B.M. 2004. The Effects of Domain Knowledge on Search Tactic Formulation. *JASIST*. 55, 3, 246–258.
- [15] Wilson, M.L. et al. 2010. From Keyword Search to Exploration: Designing Future Search Interfaces for the Web. *Foundations and Trends in Web Science*. 2, 1, 1–97.